

Metódy detekcie reči a ich využitie v telekomunikačnej praxi

Juraj Kačur*

Abstrakt

Článok sa zaoberá detekčnými algoritmi rečovej aktivity a ich možným využitím v praxi. Vysvetľuje pojem a úlohu detekcie reči a rozoberá možné úskalia s ňou spojené v tej- ktorej aplikácii. V článku sú ďalej uvedené bežne zaužívané kritéria ich klasifikácie.

Kľúčové slová : detekcia reči, hlasová aktivita, kompresia a prenos rečových signálov, rozpoznávanie reči.

Úvod

Detekčné algoritmy reči sa vo väčšine prípadov nevyskytujú ako samostatne pracujúce systémy. Zvyčajne sú súčasťou komplexných celkov, v ktorých sa podieľajú na riešení dielčích úloh spracovania reči, pričom môžu pracovať samostatne alebo v úzkej spolupráci s okolitými podsystemami. Ich použitie je implicitne dané výslednou funkciou celku a pokrýva stále väčšiu oblasť systémov spracovávajúcich reč.

1 Aplikácie detekčných algoritmov

Významnými oblasťami kde sa určite stretne s detekčnými algoritmi sú:

a) Kompresia a prenos rečových signálov

Azda najvýznamnejšou a v súčasnosti najčastejšou oblasťou, kde sa vyskytujú algoritmy detekcie rečových signálov sú systémy kompresie a prenosu reči. Tieto algoritmy totiž poskytujú významnú redukciu množstva prenášaných rečových dát. Nástupom technológií ATM (asynchronous transfer mode), internetu a aplikáciám typu „Voice over IP“, ktoré už z princípu ich činnosti nepridávajú pevné komunikačné pásmo, ich významné postavenie výrazne vzrástlo. Ved' i keď v priebehu telefónnej konverzácie musí byť zabezpečená duplexná prevádzka, je zrejmé, že za normálnych okolností je vždy jedna strana vysielajúca a druhá prijímacia. To znamená, že k úplnému využitiu oboch kanálov nedochádza. Naopak, prakticky je využitý iba jeden a to sme ešte nezahrnuli prirodzené pauzy v reči, ktoré nenesú žiadnu informáciu a preto by mohli byť tiež zanedbané. Na druhej strane treba poznamenať, že vysielajúca strana potrebuje kvôli komfortu poznať priebežné reakcie druhej strany, ktoré však zväčša nie sú rečového charakteru a na ich prenos nie je potrebná celá kapacita kanála. Z predošlej nie celkom detailnej úvahy môžeme usúdiť až na 50% redukciu dát spôsobenú len samotným detekčným algoritmom reči. Existujú dva základné prístupy prenosu rečového signálu. Systém s **prerušovaným vysielaním** počas páuz v reči vysielanie prerušuje. Kvôli väčšej prirodzenosti však prijímacia strana počas páuz generuje prijímateľovi šum pozadia. Tento prístup je veľmi rozšírený najmä v mobilnej komunikácii a stal sa aj predmetom normalizácie t.j.

* Ing. Juraj Kačur, Katedra telekomunikácií, Fakulta elektrotechniky a informatiky, Ilkovičová 3, Bratislava 812 19, Slovensko
e-mail: kacur@aladin.elf.stuba.sk, t.č: +421-7- 68279416

rečový kóder je normalizovaný vrátane detekčného algoritmu reči [4]. Druhý prístup vedie cez tzv. **variabilnú rýchlosť kódovania reči**. Namiesto prerušenia vysielania sa v neaktívnych intervaloch značne redukuje bitová rýchlosť na úkor kvality signálu, ktorá však nie je v takýchto prípadoch potrebná.

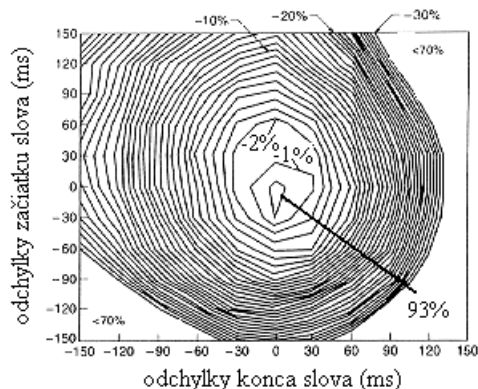
V týchto aplikáciách detekčné algoritmy musia spĺňať nasledovné požiadavky: rýchlosť detekcie, minimálne časové oneskorenie a minimálna chyba nedetekovania reči.

b) Rozpoznávanie reči

Rozpoznávanie reči je úzko spojené s nástupom moderných informačných technológií. Existujú viaceré postupy rozpoznávania reči no v súčasnosti sa môžeme skoro výhradne stretnúť s dvoma, **HMM** a **DTW**, vid' [1]. Oba prístupy v sebe zahŕňajú dve fázy: fázu tréningu a fázu rozpoznávania. Úspešný priebeh oboch fáz vplýva na výslednú kvalitu rozpoznávania celého systému. Obe tieto fázy sú však zase viac či menej závislé na presnosti určenia začiatku a konca prehovorenia. DTW (dynamic time warping) porovnáva dva vzory, referenčný a testovací a počíta medzi nimi mieru akustickej odlišnosti. Zo základného spôsobu činnosti DTW vyplýva nutnosť poznania presných hraníc oboch vyhovorení, čo je úloha algoritmov detekcie reči. Istými úpravami sa dá táto striktná požiadavka zjednodušiť za cenu väčšej výpočtovej náročnosti. Algoritmy založené na HMM (Hidden Markov model) nepotrebujú za istých okolností poznať hranice vyhovorení. Tento prístup využíva štatistické modelovanie prehovorenia, t.j. pre každú stavebnú jednotku reči, sa určí model, ktorý bude vyhodnocovať pravdepodobnosť jej výskytu pri danom vyhovorení. V prípade rozpoznávania izolovaných slov a pri niektorých krokoch tréningu modelov sú však aj tu algoritmy detekcie reči potrebné, pričom zvyšujú presnosť a znižujú časovú zložitosť celého procesu.

Na obr. č.1 je zaznamenaná závislosť úspešnosti rozpoznávania od veľkosti chyby detekcie začiatku a konca slova súčasne. Tento graf bol uvedený v [1] a týka sa rozpoznávania izolovaných slov.

Algoritmy detekcie pracujúce vo fáze rozpoznávania musia okrem presnosti spĺňať aj požiadavku na rýchlosť.



Obr. č 1. Vplyv chyby detekcie začiatku a konca slova na presnosť jeho rozpoznania.

c) Potlačenie šumu- zlepšenie kvality reči

Pri reálnych prenosoch rečových signálov, resp. už pri ich samotnom získavaní, často krát vystupuje nepriaznivý vplyv šumu. V najbežnejšom a najhoršom prípade existuje len jeden kanál (jeden zdroj poškodeného signálu), t.j. šum a signál sú už v ňom zmiešané a ani o jednom z nich nemáme explicitnú informáciu. V takomto prípade nemôžeme použiť klasické adaptívne filtre, ktoré sa úspešne používajú pri existencii dvoch kanálov s rôznym odstupom signál/ sumu [3]. Vystáva teda nutnosť odhadu charakteristík šumu zo zašumenej reči a na základe nich vhodne upraviť poškodený signál. Na ich odhad môžeme s výhodou využiť prirodzené pauzy reči čo je úloha detekčných algoritmov. Na potlačenie rušivého vplyvu aditívnych šumov známych charakteristík potom môžeme použiť napríklad nasledovné metódy: Wienerová filtrácia, rôzne modifikácie odčítania spektier, atď., vid'. [2].

Od detekčných algoritmov sa tu vyžaduje spoľahlivé označenie častí šumu, t.j. minimalizuje sa pravdepodobnosť vzniku falošného alarmu na úkor zväčšenia pravdepodobnosti nedetekovania reči. V niektorých „real time“ aplikáciách vystupuje požiadavka aj na ich rýchlosť.

d) Analýza reči – medicína, jazykoveda, atď.

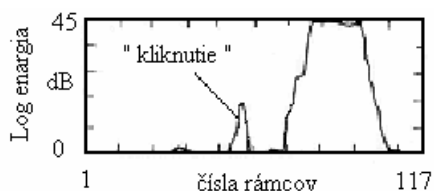
Štatisticky zdokumentované vlastnosti jazyka aj po „fyzikálnej“ stránke sú veľmi dôležité nielen pre jazykovedcov, ale aj pre medicínske účely. Časť medicíny skúma rôzne rečové chyby za účelom eliminácie ich vplyvu. Najjednoduchším spôsobom ich štúdia je práve analýza rečových signálov i keď invazívne metódy merania zväčša poskytujú spoľahlivejšie výsledky. Pri analýzach reči sa veľmi často vyskytuje potreba jej segmentácie a práve v tejto etape sa využívajú rôzne modifikácie detekčných algoritmov alebo ich častí.

Analýza zväčša nevyžaduje aby prebehla v reálnom čase, ale mala by byť čo najpresnejšia. Z toho plynie potreba nasadenia zložitejších detekčných algoritmov.

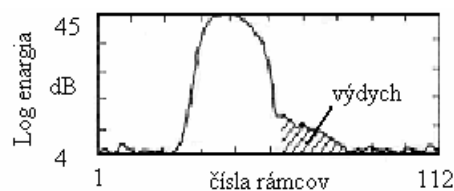
3 Problém detekcie rečových signálov

Problém detekcie reči sa dá rozdeliť do dvoch tried. Jednou je určenie hraníc slova, ktoré je vypovedané izolovane a druhou skupinou je detekovanie hraníc slov v kontinuálnej reči. Druhý problém je značne zložitejší a musí sa riešiť metódami rozpoznávania. Ďalším problémom pri detekcii je samotný charakter rečového signálu. Vyskytujú sa tu totiž samohlásky, ktoré sú charakteristické svojou veľkou energiou a rôznou dĺžkou trvania. Naopak, neznelé spoluhlásky, najmä záverové, napr. t, k, ktorých črtou je interval ticha a následne uvoľnený prúd vzduchu, sa vyznačujú veľmi nízkou energiou. Z uvedeného vyplýva, že aj v laboratórnych podmienkach je mnohokrát presné určenie hraníc slova náročné. Úloha detekcie slov sa značne skomplikuje prítomnosťou iných signálov a je tým zložitejšia, čím menší je odstup signál/ šum.

Dá sa povedať, že časové, frekvenčné ale i iné vlastnosti užitočných signálov sú len malou podmnožinou vlastností šumov. V každej oblasti sú totiž považované vždy iné signály za užitočné a všetky ostatné sú charakterizované ako šum, t.j. množiny šumov a užitočných signálov sa môžu výrazne prelínať. Šumy sú teda všetky ostatné zložky skúmaného signálu, ktoré nie sú v danom prípade nosičom požadovaného druhu informácie, ale naopak, ich prítomnosť zhoršuje možnosť získania potrebnej informácie. Za šum sú dokonca považované aj artikulačné chyby hovoriaceho, čiže samotné časti rečového signálu. Ako príklady častých artikulačných chýb môžu slúžiť tzv. puknutie pier pred vyhovorením a hlboký výdych na konci slova. Tieto situácie sú znázornené na obr. č. 2 a č. 3. Z priebehov energie v oboch prípadoch je zrejmé, že ich úspešné oddelenie od „užitočného“ signálu je pomerne ťažké a niekedy až nemožné.



Obr. č 2. Vplyv pohybu pier pred slovom



Obr. č 3. Výdych rečníka za slovom.

4 Klasifikácia detekčných algoritmov

Vzhľadom na veľkú komplexnosť danej problematiky a neexistenciu univerzálneho detekčného algoritmu, bolo navrhnutých veľa odlišných metód. Tieto sa môžu deliť podľa rôznych kritérií, a preto sú ďalej uvedené len tie najbežnejšie.

- Najčastejšie spolupracujúcimi systémami s detekčnými metódami sú systémy rozpoznávania slov a podľa toho rozlišujeme tri druhy prístupov spolupráce :
 1. **Explicitný prístup** - tento prístup je založený na tom, že detekcia hraníc slova sa uskutoční nezávisle od algoritmu rozpoznávania reči. Uskutoční sa meranie vhodných príznakov reči a na základe ich vlastností detekčný algoritmus, rozhodne o začiatku, resp. konci slova. Tieto detekčné systémy sú pomerne jednoduché a rýchle, ale dajú sa použiť iba v málo zašumenom prostredí alebo tam, kde šum vykazuje stacionárne vlastnosti.
 2. **Implicitný prístup** – tento prístup uvažuje o súčasnom procese detekcie slova a jeho rozpoznania. Neustále prebieha porovnávanie vstupných vektorov príznakov s referenčnými, ktoré zodpovedajú vzorom slov zo slovníka. Postupne sa zaznamenávajú vhodní kandidáti aj s príslušnými hranicami. Takáto metóda je dosť spoľahlivá aj v nestacionárnych šumoch. Jej veľkou nevýhodou je však značná časová a výpočtová náročnosť.
 3. **Hybridný prístup** – je to druh metód, ktoré využívajú dobré vlastnosti oboch skôr spomínaných prístupov. Dosahuje sa to predbežným stanovením skupín bodov v skúmanom signále, ktoré by mohli zodpovedať hraniciam slova, pomocou explicitného prístupu. Na presné určenie hraníc sa potom použije implicitný prístup aplikovaný na takto predbežne stanovené hranice.
- Detekčné algoritmy ďalej delíme podľa typu šumov a prostredia ich možného nasadenia, napr.: stacionárne a nestacionárne šumy, širokopásmové a úzkopásmové šumy, impulzné šumy, biely šum a farebné šumy, atď. Keďže tieto druhy šumov majú diametrálne odlišné vlastnosti, je zrejme, že na ich bezpečnú separáciu od reči musia byť použité úplne odlišné spôsoby ich detekcie.
- Ďalším významným kritériom ich klasifikácie je typ parametrov reči, na ktorom je ten - ktorý detekčný algoritmus založený. Buď sú to parametre získané z **frekvenčnej** alebo **časovej oblasti**. Tie môžu ďalej reprezentovať rôzne charakteristické vlastnosti reči ako: časovo energetické pomery, výkonové spektrum, priebeh hlasivkovej frekvencie, rôzne štatistické parametre signálu, atď.
- Ako posledné významné kritérium ich delenia môže byť rýchlosť a oneskorenie. Systémy podľa toho môžeme deliť na tzv. „on line“ alebo „off line“. Pri „off line“ aplikáciách máme k dispozícii celý signál a môžeme nad ním vykonávať operácie ľubovoľnej zložitosti. V „on line“ systémoch máme na rozhodnutie iba niekoľko predchádzajúcich vzoriek signálu, čo vyplýva z nutnosti minimalizácie oneskorenia, ktoré je v procese komunikácie dosť rušivé.

Literatúra

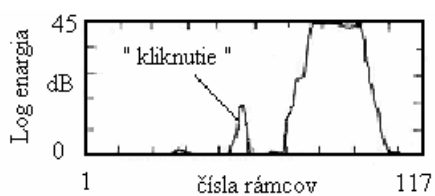
- [1] L. Rabiner, Bing-Hwang Juan : Fundamentals of speech recognition, Prentice Hall PTR, 1993

- [2] Jiří Ján: Číslicová filtrace, analýza a restaurace signálů, Vysoké učení technické v Brně, 1997
- [3] G. Rozinaj, J. Polec, J. Kotuliaková, P. Podhradský, A. Marček, S. Merchevský a kolektív : Číslicové spracovanie signálov II, FABER Bratislava, 1997
- [4] Odporúčania ETSI EN 300 973 V8.0.1, GSM 06.82

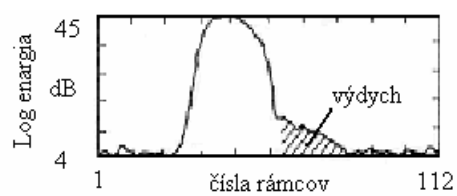
OBRAZKY



Obr. č 1. Vplyv chyby detekcie začiatku a konca slova na presnosť jeho rozpoznania.



Obr. č 2. Vplyv pohybu pier pred slovom



Obr. č 3. Výdych rečníka za slovom.